Simulation Studies of a Wide Area Health Care Network

James G. McDaniel, Ph.D., School of Health Information Science, University of Victoria, Victoria, British Columbia, Canada

There is an increasing number of efforts to install wide area health care networks. Some of these networks are being built to support several applications over a wide user base consisting primarily of medical practices, hospitals, pharmacies, medical laboratories, payors, and suppliers. Although on-line, multi-media telecommunication is desirable for some purposes such as cardiac monitoring, store-and-forward messaging is adequate for many common, high-volume applications. Laboratory test results and payment claims, for example, can be distributed using electronic messaging networks.

Several network prototypes have been constructed to determine the technical problems and to assess the effectiveness of electronic messaging in wide area health care networks. Our project, Health Link, developed prototype software that was able to use the public switched telephone network to exchange messages automatically, reliably and securely. The network could be configured to accommodate the many different traffic patterns and cost constraints of its users.

Discrete event simulations were performed on several network models. Canonical star and mesh networks, that were composed of nodes operating at steady state under equal loads, were modeled. Both topologies were found to support the throughput of a generic wide area health care network. The mean message delivery time of the mesh network was found to be less than that of the star network.

Further simulations were conducted for a realistic large-scale health care network consisting of 1,553 doctors, 26 hospitals, four medical labs, one provincial lab and one insurer. Two network topologies were investigated: one using predominantly peer-to-peer communication, the other using client-server communication. The client-server model was less expensive to operate but also less responsive to message priorities.

INTRODUCTION

Several large projects have implemented and evaluated generic wide area health care networks. These projects are the Inter-Institutional Information Exchange (3I) project [1], the Advanced Informatics in Medicine (AIM) Strategic Health Informatics Network for Europe (SHINE) [2], and others [3, 4]. As well, there are smaller prototype networks initiated by university hospitals (see for example [5, 6]). Networks are also being developed by telephone and insurance companies, for example, Ameritech [7] and Blue Cross and Blue Shield [6]. The intent of these networks, that are usually targeted at physicians, medical labs, hospitals and payors, is to support telecommunication for several diverse types of complementary health care applications.

In all but a few projects, there has been no documented effort to determine the behaviour of the proposed network and its operating cost. Although telephone companies have performed extensive analyses on the queueing behaviour of switches and telephone networks [8], little of this work has been extended to health care networks. In most cases, application network designers assume that the underlying third-party telecommunication services are adequate to support the application. This assumption can be justified as the telecommunication carriers expand their networks to meet increased demand. However, by omitting these analyses, researchers are unable to determine the dependency of network performance and cost on topology and traffic congestion.

Traffic patterns are sensitive to time of day and type of health care facility. Some institutions like hospitals and medical labs are mainly information sources. Others, like payors, are information sinks. As well, the volume of transactions varies by size and type of facility. For example, the communication traffic of a physician's office is a fraction of that of a hospital. The content of some transactions is time-sensitive. For example, the rapid delivery of lab test results is likely to be more critical than billing claims. The design criteria for a cost effective network should optimize the network configuration to meet the information exchange requirements.

This research investigated the behaviour and operating cost of a wide area health care network. It was based on actual prototype software that has been developed for the public switched telephone system. First the behaviour of mesh and star networks with symmetric loading was simulated and analyzed. Then a simulation study was performed on two models of a realistic large scale network. The results of the study were used to estimate the operating cost of the network.

BACKGROUND

One common method to estimate network performance and cost is to perform a static analysis of the network based on message characteristics, network demand and resources [9]. This type of analysis can be used to calculate an estimated response time and operating cost but the designer must make a number of assumptions about the network loading and the characteristics of network components.

A static analysis for a health care network has been described by van Lierop *et al* of the Dutch GEIN project [10]. They estimated the cost of implementing and operating an Electronic Data Interchange (EDI) network in the region of Breda, The Netherlands, consisting of three hospitals, 180 specialists and 165 GP's. They determined that one million messages would be exchanged annually. Costs were calculated based on the communication cost of a limited set of EDI transactions and the support organization needed to administer the network. The study concluded that Breda was too small to support the EDI infrastructure.

This author was unable to find other detailed analyses of wide area health care networks. It is probable that similar techniques have been used to estimate the cost of nearly all prospective networks. It is unlikely that an analysis has been conducted to determine the dynamical behaviour of a network and the cost of operating a network designed to accommodate the time-varying demands of its applications.

METHODS

Health Link [11] was a research project in which communication software was designed, developed and tested, surveys were conducted at two medical clinics, network simulations were performed and a field test was conducted for a prototype network. The simulation studies described in this paper were based on the properties of the communication software, the expected telecommunication requirements of GP's that were extracted from the surveys, and performance benchmarks that were measured in the field test.

Network Description

The *Health Link* software was designed to automatically store and forward messages using a connection-oriented protocol. It is capable of scheduling messages for pickup and delivery. Nodes can be configured to originate calls only or to both originate and accept calls. The network can be configured to support peer-to-peer communication, client-server communication and other variants. The software can use any underlying communication network that can be accessed by a serial connection but it has been designed specifically for the public switched telephone network.

A *Health Link* network is composed of one or more subnets. Depending on its routing table, a node can communicate directly with a peer in the local subnet or it can communicate with a gateway server that stores and forwards messages to the destination in the local subnet or to another gateway in a remote subnet. No message is ever transmitted more than three times from its source to its destination. The routing table for each node can be configured independently.

Every message is processed before it is sent. A Message Authentication Code (MAC) is calculated and the message is compressed and encrypted. A digital signature is generated and encrypted using the RSA public-key cryptosystem [12]. The reverse procedure is followed once a message is received. Message processing is interrupted when messages are exchanged because the MS-DOS operating system, under which the software executes, does not support preemptive multi-tasking.

Network Demand

Two clinics that participated in the field test were surveyed to determine their procedures, patient load, and the nature and volume of external communications. A messaging network would need to support the communication traffic shown in Table 1. Applications, such as making referral appointments, that are better suited for on-line communication are not shown. Messaging requirements for specialists, hospitals, medical labs, the provincial lab and the government payor were extrapolated from the messaging requirements of the GP's.

Table 1: Estimated Message Traffic for Principal Network Users

Type of Message	Source	Destination	Messages Per Day	Message Size (characters)
Medical lab test report	medical lab	GP	20/GP	175
Provincial lab test report	provincial lab	GP	3/GP	175
Emergency report	hospital	GP	2/GP	200
X-ray report	hospital	GP	2/GP	300
Lab test report	hospital	GP	2/GP	250
Operating room (OR) report	hospital	GP	2/GP	1000
OR booking	GP	hospital	2/GP	175
OR booking confirmation	hospital	GP	2/GP	175
Outpatient clinics	GP	hospital	2/GP	175
Specialist referral	GP	specialist	3/GP	1000
Specialist referral	specialist	specialist	3/source	1000
Specialist report	specialist	GP	5/specialist	200
Health insurance claim	GP	government	1/GP	2000
Insurance claim confirmation	government	GP	1/GP	100
Medical record transfer	GP	GP	3/GP	1000
Medical record transfer	hospital	hospital	10/destination	500

Discrete Event Simulation Modeling
Discrete event simulation models [13] were constructed for the nodes and gateway servers. Simulation transactions, that were computational analogs of messages, were subjected to equivalent processing and queueing delays.

Canonical Steady State Models: Symmetric mesh and star node clusters were simulated using a program that exploited cluster symmetry through a multi-nodal driver. Each node in a cluster was assumed to have a single communication port. For star clusters, network performance was analyzed for a varying number of gateway ports. The number of messages exchanged during the simulations were nominally 3,000 messages and 1,000 messages for the mesh and star clusters, respectively. The results of the discrete event simulations were validated by performing analytic studies [14] on the same models.

Large-Scale Dynamical Models: Simulations were based on a realistic asymmetric model describing the health care system in the Province of Saskatchewan, Canada. In this analysis, a program was written that used a rule-based object-oriented approach to construct the network topology. Events such as message creation and connection initiation were determined by random events constrained by time schedules and node configurations. Eight subnets were defined for a network of 1,585 nodes (see Table 2). In the course of a simulated day, approximately 80,000 data messages were exchanged. Except for those data messages from the provincial lab, all messages were acknowledged by the destination node upon receipt.

Analyses of two models were performed. One was based on peer-to-peer communication among the physicians, hospitals and gateway servers within a given subnet. This model configured those physicians who were in the same subnet as one or more labs, to poll those labs six times daily. In cases where a lab delivered data messages outside its subnet, it forwarded the messages at medium priority to the gateway server without delay.

The second model was based on client-server communication within each subnet. Each node within a subnet delivered messages to its gateway server and polled the gateway server for messages. Polls were scheduled hourly. Gateway servers exchanged messages among subnets without delay.

Model Parameters: Table 3 shows some of the timing parameters used by the simulation models. All communication links were 2,400 bits per second serial connections. The base processor at each node was assumed to be an Intel 80386SX25. The number of serial ports at a node determined the CPU factor that was used in calculating the message processing times. For example, a computer with two ports was assumed to have twice the CPU power of the 80386SX25.

Table 2: Composition of the The Saskatchewan Model

Region	GP's	Specialists	Hospitals Municipal	Hospitals Regional	Medical Labs Private	Medical Labs Provincial	Insurer
Moose Jaw	46	8	0	2	0		
North Battleford	92	16	0	4	0		
Prince Albert	92	16	0	4	0		
Regina	245	185	4	0	2	1	1
Saskatoon	338	299	4	0	2		
Swift Current	46	8	0	2	0		
Weyburn	46	8	0	2	0		
Yorkton	92	16	0	4	0		
Total Nodes	997	556	8	18	4	1	1

Message priority was implemented by using priority holding delays. Messages were collected in a mailbag that was held for the least priority delay of any one of the enclosed messages. A connection was initiated when the priority delay expired or when the receiver initiated a connection. For the steady state simulations, high, medium and low priority delays were zero, three and five minutes, respectively. For the dynamical simulations, these parameters were zero minutes, three minutes and four hours, respectively. Data messages were exchanged at either high or medium priority. Acknowledgment messages were exchanged at low priority.

The payload of data messages was fixed at 1,000 characters for the steady state simulations. The mean payload of messages sent in the dynamical simulations matched the message sizes shown in Table 1. Acknowledgment messages carried no payload.

Calibration: The timing parameters shown in Table 3 were measured in benchmark tests. The effective data exchange rates were taken from a stress test conducted over a period of three months at four different sites. The stress test was run using an automatic driver at each node that sent messages timed according to a Poisson distribution. The stress test analysis produced a performance curve for an actual four node cluster. For each simulation program, the performance of a modeled four node cluster was matched to the actual cluster.

Model Initialization: Two methods were used to compensate for the bias caused by starting the simulations from an empty and idle state in which there were no messages. For the steady state simulations, swamping [15] reduced the initialization bias to less than two percent on the mean message delivery time. For the dynamical simulations, a warm-up period of one simulated day removed the bias.

Table 3: Simulation Parameters

Timing Parameter	Value
Dial time (sec)	15
Ring and connect time (sec)	12
Relisten time (sec)	14
Modem reset time (sec)	10
Line idle detection time (sec)	60
Effective full-duplex data rate (bps)	362
Effective half-duplex data rate (bps)	770
Minimum wait between retries (sec)	180
Fixed protocol overhead (sec/msg)	2
Message processing time (sec/msg)	21

Calculating Standard Error: Simulations were conducted in a series of independent replications with random starting points to estimate the standard error of various performance measurements [13].

RESULTS

The results of the two series of simulation studies are given in turn. The simulations of the steady-state models provided insight into the behaviour of a network where all nodes were similar. The simulations of the dynamical models explored the behaviour of similar topologies in a realistic environment where the nodes experienced stochastically cyclical, asymmetric loads that characterize the health care system.

Canonical Steady State Models: The studies performed on the mesh and star node clusters indicated that both topologies are capable of supporting the anticipated mean load of 50 messages per day per node. Mesh clusters with up to 1,000 nodes would have an operating region from zero to 12 messages per hour per node. The cluster exhibits uniform performance in this region, achieving end-to-end delivery of medium priority messages, each having a data payload of 1,000 characters, in under 8 minutes.

Changes to the priority delay can affect the throughput as well as the performance of the mesh cluster. For small clusters, the number of messages exchanged during a single connection increases with the priority delay thus increasing port efficiency. The impact of the priority delay on throughput decreases as the number of nodes in the cluster increases.

The performance and throughput of star clusters depend on the capacity of the gateway server. The cluster exhibits long end-to-end message delivery time for both light and heavy network loads. When the cluster is lightly loaded, nodes rarely connect with the gateway server to send and receive their messages. When the cluster is heavily loaded, the gateway server is under contention. A star cluster is able to exceed the throughput of a mesh cluster if the server is configured with a sufficient number of ports. For example, a throughput of 15 messages per hour per node can be achieved in a 1,000-node cluster with 250 server ports. This configuration results in a mean end-to-end message delivery time of one hour.

Priority holding delays have little beneficial effect in star clusters. Increasing the delay has the effect of improving the throughput of the sender but there is no benefit to the receiver. Messages are held by the gateway server for pickup regardless of their priority. Priority delays have a negative impact on performance since they increase the message holding time at their source.

Two general observations were also made:
1) message processing requires as much time as does communication and 2) if an acknowledgment message is sent to confirm the delivery of each data message, the cluster throughput is reduced by nearly 40 percent.

Large-Scale Dynamical Models: Ideally, the models chosen to represent a network in the Province of Saskatchewan, Canada would optimize performance and operating cost. Moreover, we were interested in exploring the differences between peer-to-peer and client-server communication. However, optimization of these two models is difficult because of the many configuration and scheduling parameters. Examining the results is equally problematic as there are many interesting outcome measurements. A few of the more prominent observations follow.

The analyses of the peer-to-peer and client-server models indicate that both topologies can be used for a wide area health care network. The two models differ with respect to performance and operating cost. In order to make the mean end-to-end message delivery times comparable between the two models, each node in the client-server model was scheduled to poll its gateway server at a mean rate of once per hour. This resulted in mean end-to-end data message delivery times of one hour, 23 minutes for the peer-to-peer model and one hour, 24 minutes for the client-server model.

The peer-to-peer model exhibited the same sensitivity to priority delays as did the mesh model. Figure 1 shows two cumulative density functions for medium priority messages sent among physicians, one for messages exchanged among subnets and one for messages exchanged within the same subnet. Figure 2 shows the same functions for the client-server model. Table 4 gives a comparison between the two figures. The delivery time of messages exchanged in the client-server model was almost ten times that of messages in the peer-to-peer model; a consequence of the hourly polling frequency by the destination nodes in the client-server model.

The telecommunication cost of the peer-to-peer and client-server models was evaluated. The telephone and Datapac (a public packet switched network) tariffs were applied to the traffic patterns and volumes derived from the simulation models. The more economical service, voice telephone or packet switched communication, was chosen for each node. Table 5 shows a comparison of telecommunication monthly basic access and usage charges before applying several cost-optimizing strategies.

Figure 1: Mean End-To-End Delivery Time for Medium Priority Messages Sent Among Physicians in the Peer-To-Peer Model

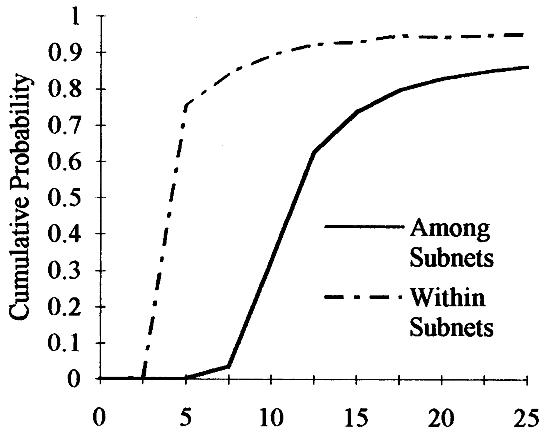
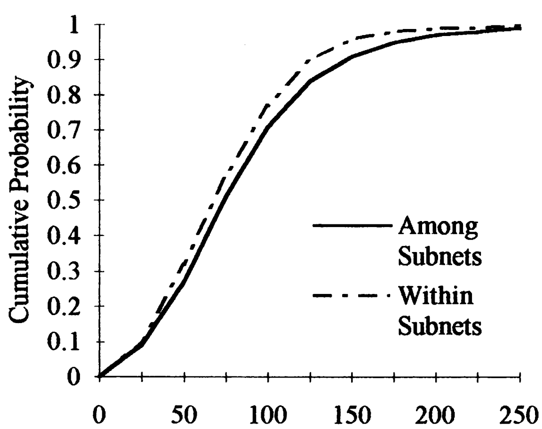


Figure 2: Mean End-To-End Delivery Time for Medium Priority Messages Sent Among Physicians in the Client-Server Model



The peer-to-peer model relied heavily on telephone communication. Each node was supplied with at least one dedicated communication line, because every node in this model was configured to accept calls. This is not a prerequisite as there are devices available that sense the communication mode of incoming calls. The peer-to-peer model incurred long distance telephone charges within each subnet. Although ideally a subnet would be defined within single local dialing area, Saskatchewan has a widely dispersed population that is 37 percent rural.

Table 4: A Comparison of Medium Priority Messages Sent Among Physicians

Simulation Model	Among Subnets	Within Subnets
Peer-To-Peer		
Sample size	5,361	8,845
Median delivery time	11:40	4:10
Mean delivery time	22:59	9:03
Client-Server		
Sample size	5,537	9,064
Median delivery time	1:17:30	1:11:40
Mean delivery time	1:26:44	1:16:59

Table 5: A Comparison of Monthly Telecommunication Charges ($Cdn)

Telecommunication charges	Peer-To-Peer	Client-Server
Telephone access	57,722	6,653
Datapac access	2,420	3,563
Long distance telephone	77,293	0
Datapac usage	7,849	52,050
Datapac volume discount	-706	-4,029
Total charges	144,578	58,237
Charge per node	91.22	36.74

For the client-server model, because all call connections terminated at the gateway server, it was possible to use public Datapac ports. Universal Datapac Access (UDA) was used which included long distance telephone charges in the total Datapac usage charges.

The telecommunication cost of both models can be reduced by tuning call connection schedules. A saving of approximately five dollars per month can be made without significantly changing the behaviour of either model. Other gains may be realized by configuring hybrid networks that are better suited to the traffic patterns and the tariff schedules.

CONCLUSION

Two series of simulation studies were conducted for a telephone-based wide area network. Canonical mesh and star model networks were examined. The mesh network was more sensitive to message priority than the star network. The mesh network also was able to deliver messages in a shorter time than the star network. The star network could be configured to have a greater throughput than the mesh network. Both networks are easily capable of meeting the projected demands of a wide area health care network.

Simulation models with two topologies for a large-scale wide area health care network were constructed. These studies indicated that both peer-to-peer and client-server topologies are practical. The simulation of the peer-to-peer topology predicted that message delivery can be expedited by setting message priority but at a telecommunication cost that is greater than that of the client-server topology. The telecommunication cost in Canadian dollars of a network delivering 1,100 messages per month per node was predicted to range from $32 to $87 per month per node. This is a per-message cost of between $0.03 and $0.08.

The communication networks that were modeled by these studies rely on low speed telephone and packet switched technologies that are accessible to nearly all users. Although these technologies do not allow for high speed broadband communication that is needed for transmitting large messages such as digital radiographic images, it does support the transmission of smaller messages that are more commonly exchanged by health care providers.

The simulation techniques used in this research can be extended to other geographic regions, other applications and other telecommunication technologies. Simulations such as these could predict the behaviour and operating cost of a network before undertaking its costly installation.

ACKNOWLEDGMENTS

Support for this research was given in part by the Science Council of British Columbia under HDF #43-89, MART #90/91-11, HDF #7-91, and Technology B.C. #179 (T-3) Grants. Much of the financial support for my studies has been given by the University of Victoria and the Medical Research Council of Canada.

Reference

[1] Hasman, A., Arnout, P.C. and Boon, P. The 3I Project. In: *Telematics in Medicine*, Duisterhout, J.S., Hasman, A. and Salamon, R. (eds.). Amsterdam: North-Holland, 1991: 163-173.

[2] Commission of the European Communities DG XIII Information Technologies and Industries, and Telecommunications. AIM 1993. Annual Technical Report on RTD: Health Care. 1993: 181-186.

[3] Duisterhout, J.S., Hasman, A. and Salamon, R. (eds.). *Telematics in Medicine*. Amsterdam: North-Holland, 1991.

[4] Science Council of British Columbia. *Sharing Health Information: An Overview of Fifty Projects*. SPARK Health Sector Report. Burnaby: Science Council of British Columbia, 1992.

[5] Bahensky, J. University's Statewide Network Links Rural Doctors. *Healthcare Informatics* 1992, March: 10-12.

[6] The Impact of EDI & ANSI Standards on Administrative Cost Containment. *Healthcare Informatics* 1992, November 9(11): 90-96.

[7] Ameritech advertising brochure, Chicago, circa 1993.

[8] Briley, B.E. *Introduction to Telephone Switching*. Reading: Addison-Wesley Pub. Co., 1983.

[9] Ellis, R.L. *Designing Data Networks.* New Jersey: Prentice Hall, 1986.

[10] van Lierop, D. et al Feasibility of a regional EDI-network in Dutch health care. In: *MedInfo 92*, Lun K C et al. (eds). Amsterdam: North-Holland, 1992: 89-93.

[11] McDaniel, J.G. *Health Link: A Wide Area Telecommunication Network for Health Care Providers.* Unpublished Ph.D. Dissertation. University of Victoria, Canada, 1994.

[12] Rivest, R. L., Shamir, A. and Adelman, L. A Method of Obtaining Digital Signatures and Public-Key Cryptosystems. *Communications of the ACM* 1978,2: 120-126.

[13] Banks, J. and Carson, J.S. II. *Discrete-Event System Simulation*. Englewood Cliffs: Prentice-Hall, 1984.

[14] Kleinrock, L. *Queueing Systems, Volume I: Theory*. New York: John Wiley & Sons, 1975.

[15] Banks, J., Carson, J. S. II and Sy, J. N. *Getting Started with GPSS/H*. Annandale: Wolverine Software Corp., 1989.